

適用於物件導向應用之視訊分割研究

研究生：詹益鎬

指導教授：林大衛 博士

國立交通大學

電子工程學系暨電子研究所



實現物件化導向 (object-oriented) 視訊應用之關鍵在於穩健的視訊分割 (video segmentation) 技術，本論文研究自動視訊分割技術。我們提出數個自動分割的方法，最後並使用一個簡單的方法來將分割後的視訊合成新的畫面，以檢視其效能。

首先，我們提出一種區域分割 (region-based segmentation) 和移動分析的方法，此研究之重點在於如何啟始區域集合、降低過小區域數量、將區域域集整合使物件符合認知、及追蹤變形物件。為了易於實現的理由，我們選用區域成長 (region growing) 方法以獲得影像的分割區域集。為使該方法適用於自然影像，我們作了一些修正使接著而來的影像得以繼承住區域集。由於區域集在開始時仍屬粗略，特別是當區域隨著時間具有較明顯變化時則必需重新分割；該區域發生較明顯變化的主因在於區域移動，而估測區域移動的過程通常是最為耗時的，為此我們提出一種簡化的移動估測方法，最後，不以高階人工智慧作預判，我們根據各區域移動向量的訊息，整合移動相似度高的區域達成物件萃取的目的。為呈現萃取物件之效能，我們分別在主觀及客觀上作評估。

其次，我們提出一個可自動分割重疊物件的演算法。該演算法使用低層級的時空信號處理，透過分析畫面中所呈現的運動、畫面間像素值的變化、邊緣資訊、及影像區域中紋理的一致性，來進行分割。演算法的設計，是希望能具有以下幾項功能：(1) 能分開多個互相重疊的物件，其中各物件可做複雜或相當快速的運動；(2) 能處理物件的變形；以及 (3) 能處理物件的出現與消失。本演算法的一個特出之處為其多階層的架構。其第一階層(即最低層)從每張畫面中萃取出了一組低層級的特徵。第二階層(即中間層)利用這些特徵，將各畫面分割為一個移動部分(稱為前景, foreground)和一個靜止部分(稱為背景, background)。第三階層(即最高層)則是在前景中，根據運動分析及型態處理(morphological operations)，鑑別出其中互相重疊之物件，並作各物件的追蹤。經使用一些不同特性的視訊做實驗的結果，顯示本演算法獲得的物件邊界具有相當程度的準確性。

第三，在自動分割領域中，一個關鍵問題在於如何準確的判定物件的邊界。我們使用一種銳緣處理(edge-linking)方式將邊界以分段考慮(segment-based consideration)的方式去準確銳化已分割出來的移動物件邊界，與此相類似的文獻目前仍相當稀少，我們提出的設計著眼於效率上的改進及較高的準確性。方法上，首先是以一種離線的方式粗略的偵測出物件區域，例如，先以變動偵測後再加以緊實物件區域內部，這部分程序稱之為遮罩草繪(mask sketch)，最後，找出最外緣邊界且根據連接性將其分段，並以搜尋最短路徑演算法(shortest-path algorithm)將各段分別銳化。在此想法上的特色是：降低遮罩草繪的複雜度及更有效的降低搜尋區域，實驗顯示，我們的方法能產生很好的效果，為了能將萃取出之物件在之後的畫面上穩健取得，我們還使用了雙向移動估測的方式追蹤並且重複的使用上述銳化方式。

第四，為使銳緣處理達到即時，我們設計一個以圖素為基礎的方法。此演算法中最主要的創新之處在於一個依據運動狀況來追蹤可變形物件的方法以及一個稱為遮罩精煉(mask refinement)的功能方塊。於前者中，我們兼用前向與反向的運動估計以求穩健，但其運算複雜度並不高。於後者中，我們採用由合適之銳緣偵測器所尋得的銳緣，加以型態處理(morphological edge-oriented processing)以刻劃出物件邊界。我們提供一些實驗結果以顯示此演算法的效能。這些結果顯示。除了背景有劇烈運動及運動物件與背景之對比相當低的視訊，此演算法的效果都相當良好。其所需之計算時間也顯示適合於桌上型或可攜型之多媒體應用。

最後，萃取出之物件將更進一步應用於畫面組合上，為此用途，我們描述並評估提出的物件的放大縮及小方法，它們就像是繪圖板上的工具，當編輯各種圖案時使用多層的安排將各種固定的影像及分割出的視訊物件組合起來，以豐富

視訊應用。從視訊編碼的觀點來看，提出的放大和縮小方法可用於可調空間編碼（spatial-scalable coding）技術上。



Research in Video Segmentation Techniques for Object-Oriented Applications

Student: Yih-Haw Jan

Advisor: Dr. David W. Lin

Department of Electronics Engineering
& Institute of Electronics
National Chiao Tung University

Abstract

The robustness and quality of video segmentation has been one of the main reasons that the standards of the object-oriented function and related applications have not been able to be implemented. This thesis proposes automatic video segmentation schemes. Finally, we use a simple method by composing scenes to survey performances.

First, this study develops an automatic method that combines region-based segmentation and motion analysis. This method focuses on how to begin to obtain a region set, how to eliminate too small regions, how to integrate high similarity regions, and how to maintain similarities during processing. For simplicity, this study uses the region-growing method to segment regions. Furthermore, his method is adjusted for application to natural images. The update of this method for use in subsequent images segmentation works well to hold the region set. The update modifies the segmentation by setting the geographical homogeneities and adjusting the object boundaries when the original image changes significantly in latter frames. Regional movement results from the need for update. However the method frequently takes the highest complexity in video-making processes. This study proposes an easy method of simplifying the complexity. Without higher-level intelligence, the regions are integrated for object extraction while the regions exhibit similar motions, and the tracking efficiency is also presented.

Second, this work presents an algorithm for automatic segmenting overlaid video objects. The algorithm employs low-level spatio-temporal signal processing and the segmentation is based on the analysis of apparent motion, inter-frame pixel value changes, edges, and textural homogeneity of image regions. The algorithm is designed to separate multiple overlaid objects and enable them to do complicated or relatively fast movements, to handle object deformation, and address object appearance and disappearance. The above functions distinguish the algorithm from some other recently published ones. The algorithm has a three-tier structure. The first and lowest tier extracts a set of low-level features from each video frame. Meanwhile, the second and middle tiers employ these features to segment each video frame into a moving part (called the “foreground”) and a stationary part (called the “background”). The third and highest tier then identifies and tracks the overlaid objects in the foreground via motion analysis and morphological operations. Experiments involving several different videos show that the algorithm can yield reasonably good identification of object boundaries.

Third, a critical problem in this area is how to accurately identify the object boundaries. This work studies the edge-linking approach. Restated this study uses segment-based consideration to accurately locate the boundaries of the moving objects in video segmentation. Similar existing methods are quite rare. This investigation devises a scheme designed for efficiency and enhanced accuracy. The proposed scheme first obtains a rough outline of an object via a suitable method, such as, change detection. The scheme then obtains a relatively compact image region that contains the object, via a procedure termed “mask sketch.” Finally, the outermost edges in the region are identified and linked using a shortest-path algorithm. The ideas employed in mask sketch effectively reduce the search area. Experiments show that the scheme achieves good performance. Obtained objects iteratively keep employing the same approach accurately determining object boundaries and bi-directional motion estimation for robust tracking.

Fourth, we study an approach to update object border of the moving object for real-time purpose. This work presents a pixel-based algorithm for achieving this goal. The proposed algorithm employs change detection and motion estimation to identify and approximately track de-formable moving objects. The main novelty of the algorithm comprises a function block termed mask refinement, which performs morphological edge-oriented processing to accurately delineate the object boundaries, using the edges located using a suitable edge detector. The motion estimation is object-based to ensure robustness in object tracking, both forward and backward

motion estimations are conducted, and the method has relatively low complexity. This investigation presents experimental results to illustrate the subjective segmentation performance of the algorithm. The computational time required by the algorithm is considered appropriate for real-time desktop or portable multimedia applications.

Finally, extracted video objects are further applied to scene composition. This study describes and a method for enlarging or shrinking objects for this purpose. Like a palette, editing various patterns via multi-layer arrangement with fixed images and segmented video objects varies in rich video applications. Moreover, from a video coding perspective, the enlargement and shrinkage methods are amenable to spatial-scalable coding.



誌謝

這篇論文的完成，首先要感謝我的指導老師林大衛教授多年的教誨與指導。尤其老師的研究精神與待人處事的態度，為我樹立了一個良好的典範。此外，每當我的研究過至叮瓶頸時，老師總是不厭其煩的指引我正確的方向。在此我要再次表示我十二萬分的謝意與敬意給林大衛教授。

我也要感謝通信實驗室的師長們，他們提供了優良的研究環境及設備，使我的研究能順利的進行。另外，感謝諸位學長、同學、學弟妹們，在彼此互相的討論、砥礪下，使我獲益良多。

當然，我要獻上最深摯的感謝給我的父母與家人：最敬愛的爸爸、媽媽，最親愛的兄弟姐妹們，因為有你們在背後默默的支持與鼓勵，才使我得以堅持到最後，順利完成學業。願與你們一同分享這份喜悅。

